

# Designing of Information System for Semantic Analysis and Classification of Issues in Service Desk System

Ksenia Lokhacheva<sup>a</sup>, Denis Parfenov<sup>a</sup> and Maria Lapina<sup>b</sup>

<sup>a</sup> Orenburg State University, Prospekt Pobedy, 13, Orenburg, 460018, Russia

<sup>b</sup> North-Caucasus Federal University, Pushkin St., 1, Stavropol, 355017, Russia

## Abstract

The paper describes the designing of Information System for Semantic Analysis and Classification of Issues in Service Desk System. The concept of a Service Desk system and problems of its using are described; several mathematical models and methods of text analysis and text classification are studied; an analysis of system usage options, construction of a system scheme and a Class diagram were held.

## Keywords<sup>1</sup>

reinforcement learning, machine learning, algorithmic trading, market make, market liquidity

## 1. Introduction

Text classification (a.k.a. text tagging or text categorization) is the process of automatic text analysis and assigning a set of pre-defined tags or categories based on text content. Automatic text classification can be based on production models with pre-defined strategies, or on machine learning methods. The problem of classifying user requests (issues) in the Service Desk system is one of the examples of text classification tasks.

Service Desk is a system that manages incidents (service disruptions) and service requests (routine service related tasks) along with handling user communications for things like outages and planned changes to services. The responsibility of the Service Desk operator is to handle such requests: he gives instructions for troubleshooting or fixes them on his own via remote access. In order to effectively solve the problem, the operator must correctly classify it, that is, determine the area to which this issue belongs. In this case, the operator can:

- 1) Use the Data Base;
- 2) Use the built-in Service Desk search (for similar user issues);
- 3) Classify issue on his own according to his professional experience;
- 4) Search in the Internet (Google, Yandex, etc.).

The main difficulty of the content searching for issue in Service Desk systems is that characteristics of the same class issues are described in different words. In addition, descriptions often contain slang, grammatical errors and forms of mailing, since most of the requests are received by e-mail. All these factors significantly complicate the process of issues classification, and therefore increase the duration of finding the correct matches and solutions for issues.

---

YRID-2020: International Workshop on Data Mining and Knowledge Engineering, October 15-16, 2020, Stavropol, Russia  
EMAIL: ksenia.lohacheva.97@mail.ru (Ksenia Lokhacheva); parfenovdi@mail.ru (Denis Parfenov); norra7@yandex.ru (Maria Lapina);  
ORCID: 0000-0002-8073-0710 (Ksenia Lokhacheva); 0000-0002-1146-1270 (Denis Parfenov); 0000-0001-8117-9142 (Maria Lapina);



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Related works

Most companies in one way or another work with clients and provide user support service. In addition, technical support of internal processes is a question of great importance for successful company management.

In work [3] negative aspects of the wrong organization of Technical Support Department work are described, namely:

- lack of fixed areas of competence creating a misunderstanding of the importance of the functions performed;
- risk of the particular user request loss in the total amount of requests and managers' orders as a result of an unregulated request form;
- high dependence of the company's work on the "key" specialist, which occurs when a certain type of work is regularly performed by one employee.

Service Desk systems are able to ensure high-quality interaction between all members of the business process. The main tasks of Service Desk systems are the receiving and processing requests, i.e. the client creates request (ticket) and Service operators process it. With the use of the Service Desk system, it is possible to improve the work of all Service operators of the company.

Processes in the Service Desk systems regulate all the difficulties that arise in the work of the IT-Department [4]:

- Incident Management
- Problem Management
- Change Management
- Release Management
- Service Level Management
- Financial Management
- Availability Management
- Capacity Management
- Continuity Management
- Information Security Management

Thus, according to the described functions and tasks of the Service Desk system, automation of some processes using semantic analysis and requests classification in order to predict the most likely solution to the problem without additional involvement of specialists seems relevant.

Natural languages texts analysis involves two stages:

1. word embedding, that includes Parsing, Part-of-speech tagging, excluding stop-words, digits, Stemming (or Lemmatization).
2. model training on pre-labeled data and text classification.

Due to the fact that automatic processing of text information is becoming more and more relevant and in demand, nowadays there is a large number of studies on methods of models training.

In [1] and [2], a comparative analysis of text classification methods is carried out. Both papers present a formal formulation of the text classification problem, describe classification methods, and provide a comparative analysis of classifier training methods using machine learning technologies, including the Bayes method, k-nearest neighbors algorithm, least squares method, support vector machine, and methods based on artificial neural networks. The main criteria for evaluating the quality of the classification were a combination of precision and recall. Based on the study [1], it was concluded that the best ratio of these characteristics is achieved using the methods of support vector machine and convolutional neural network. At the same time, the speed of the Bayes method is one of the highest, but the accuracy for different experiments varies. According to the study [2], the least squares method showed the best results in terms of recall, while the support vector method was the best in terms of precision. A comparative analysis of the considered classification methods based on studies [1] and [2] is presented in table 1.

**Table 1**

Analysis of the effectiveness of text classification methods

Method	Average precision, p	Average recall, r
Support vector machine	91,3%	84,2%
k-nearest neighbors algorithm	79,5%	77,1%
Convolution neural network	91,5%	81,3%
Bayes method	74,6%	82,5%

One of the most popular directions in text analysis is word embedding – techniques in natural language processing for mapping words to a vectors of real numbers. The article [5] describes the concept of such mapping, shows the word2vec model – one of the most popular models at the moment, explains how one can use the word2vec implementation in Python for classification tasks (user determination on the visited sites sequence) and regression tasks (predicting the popularity of articles).

Several papers are devoted to the description of full-fledged systems for the analysis of text documents and subsequent document processing. Thus, the article [6] describes the Stanford CoreNLP framework, which provides opportunities for processing texts in natural languages. The framework is used under JVM, and provides most of the main stages of natural language processing from tokenization to coreference resolution. The article describes the system architecture, simple templates for using framework components, and the principles of annotation (adding metadata to code) using Stanford CoreNLP. The framework supports English, French, German, Chinese, and Arabic, but does not support Russian.

Another article [2] describes a software complex that includes an open system for automated text analysis, an “Automated text analysis” portal, and a set of services and mobile applications. The paper presents the architecture of a tool set for automated text analysis in Russian, that links a data management system, an open system for automatic text analysis, an Internet portal, and a mobile application. The paper does not present a study of the accuracy, reliability and adequacy of the proposed solution. In addition, it is not possible to classify texts that have a size comparable to the size of user requests in Service Desk systems.

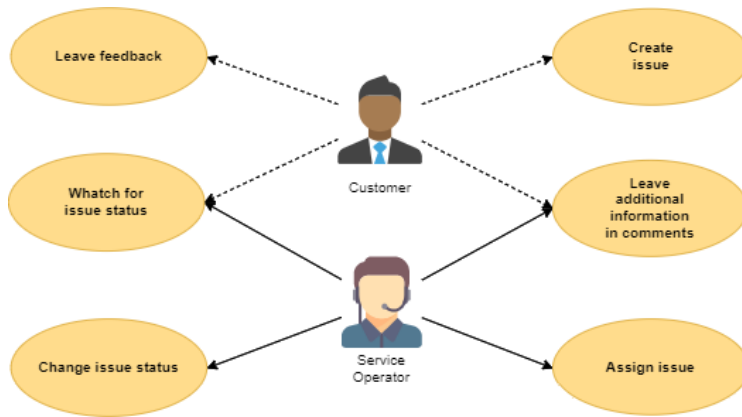
Based on the relevance of the problem and the lack of adequate analogues that support Russian, it was decided to develop an Information System for Semantic Analysis and Classification of Issues in Service Desk System.

### 3. Problem statement

The goal is to design an information system for semantic analysis and classification of issues in Service Desk system. Typically, Service Desk systems support a three-level client-server architecture, in which the client (user interface), application (hardware and software), and data (DB and DBMS) levels are physically separated.

The following options are available for using the Service Desk system (figure 1).

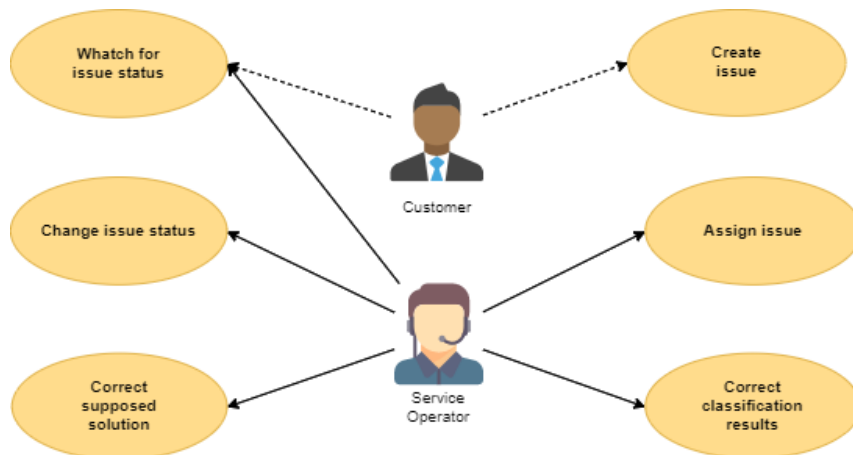
We suppose that each request left in the Service Desk system will be pre-processed before it is included in the list of requests to be executed. At the same time, the pre-processing will consist of semantic analysis of semi-structured data extracted from the particular issue, classification of the issue (searching for the most appropriate executing Department (or team) in Technical Support Department), and selection of a possible solution based on the analysis of solutions of previously closed issues of the same category.



**Figure 1:** Variants of Service Desk system usage

After pre-processing, the request is added to the list of requests to be executed for a specific Department. Employees of this Department can assign any request to themselves. If, after the first issue reviewing, the technical service operator agrees with the results of the classification, he can review a possible solution, try to apply it, and then, if the solution offered by the system did not help, note this fact in the issue description and offer a new one. If at some point of issue execution it becomes clear that the classification was incorrect, the technical service operator can detach this issue from himself and move it to the list of general open issues. After executing and closing issues from the list of general open issues, an employee who executed it must leave appropriate comments on the task (about the executing Department and the correct solution).

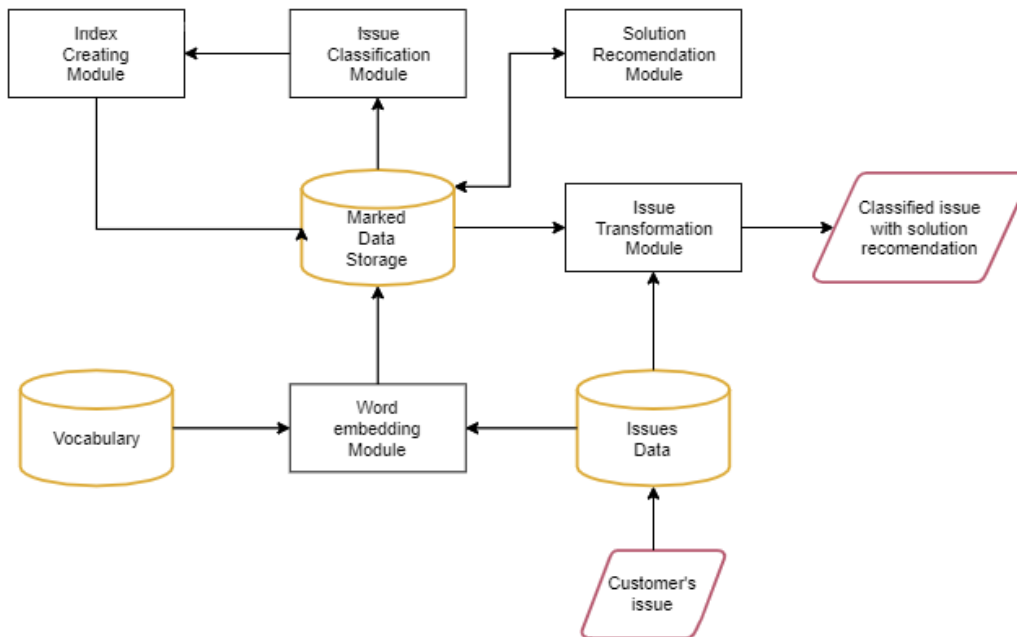
As a result, the options for interacting with the proposed system look as shown in figure 2.



**Figure 2:** Variants of interacting with the proposed system

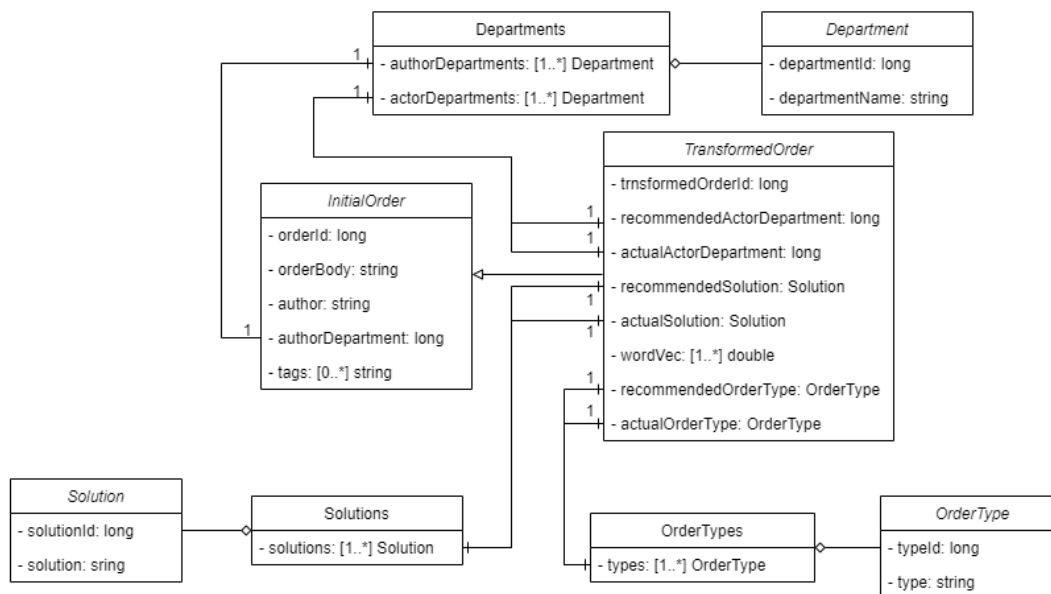
#### 4. System design

The scheme of the developing system is shown in figure 3. In this case, Issues Data, Vocabulary, and Marked Data Storage are components of the Data Storage.



**Figure 3:** Variants of interacting with the proposed system

The user leaves the request in the Service Desk system, its data is stored in the Issues Data storage, then the entire request is vectorized using vocabularies (databases) of the Russian language. The marked data is sent to the appropriate storage, where the Issue Classification Module pulls it up. After classification the index of the current issue to update information in the Marked Data Storage is held. In addition, after issue classification, a possible solution should be proposed. As the output, the system converts the original request, adding the assigned task class, the executing Department, and the possible solution for the issue.



**Figure 4:** Variants of interacting with the proposed system

Figure 4 shows a Class diagram of the main entities of the designing system. In this case, the entities are:

- Department, responsible for the Department of the company where employees work. The DTO “Departments” is linked to the Department entity by an aggregation relationship. It stores information about all departments of the company where employees who creates issues (the “authorDepartments” attribute) and employees who execute issues (the “actorDepartments” attribute) work.

- InitialOrder, responsible for initial information of received issue. This entity contains the following attributes: the issue identification number (orderId), the issue body (orderBody), information about the issue author(author), information about the Department where the issue author works (authorDepartment), in this regard, this entity is linked with the “Departments” DTO by an association relationship, and a list of tags that the author could add to the issue description to specify the problem (tags).

- TransformedOrder, responsible for information about the transformed request. This entity inherits the attributes of the InitialOrder entity and also has:

- a) the transformed issue identification number (newOrderId);

- b) vector representation of the issue body (wordVec);

- c) the system-selected request type (class) (recomendedOrderType), the actual request type (class) (actualOrderType), these attributes link the TransformedOrder entity to the “OrderTypes” DTO;

- d) the system-selected Department whose employees could solve the issue (recomendedActorDepartment), the actual Department whose employees solved the issue (actualActorDepartment), these attributes associate the TransformedOrder entity with the “Departments” DTO;

- e) the system-selected issue solution (recomendedSolution), the actual issue solution (actualSolution), these attributes link the TransformedOrder entity to the “Solutions” DTO.

- OrderType, responsible for the classification type of the issue. The “OrderTypes” DTO is associated with the OrderType entity by an aggregation relationship, and it stores information about all possible order types (the “types” attribute).

- Solution, responsible for the type of issue solution. The “Solutions: DTO is associated with the Solution entity by an aggregation relationship, and it stores information about all possible types of solution requests (the “solutions” attribute).

These entities are the main components of Issues Data and the Marked Data Storage (figure 3).

The designing system will be implemented as a plug-in for one of the most famous Service desk systems – Jira.

## 5. Conclusion

The paper describes the designing of Information System for Semantic Analysis and Classification of Issues in Service Desk System. The following points are mentioned:

1. the concept of a Service Desk system and problems of its using are described;
2. several mathematical models and methods of text analysis and text classification are studied;
3. the information system for semantic analysis and classification of issues in Service Desk system was designed, an analysis of system usage options, construction of a system scheme and a Class diagram were held.

To implement this system, further research of vectorization methods, classification methods, and solution recommendations methods that are compatible with the Atlassian SDK are necessary.

## 6. Acknowledgments

The study was carried out with the financial support the grant from the President of the Russian Federation for state support of leading scientific schools of the Russian Federation (NSH-2502.2020.9).

## 7. References

- [1] A.I. Kadhim "Survey on supervised machine learning techniques for automatic text classification." *Artificial Intelligence Review* 52.1 (2019): 273-292.

- [2] A.K. Abasi, A.T. Khader, M.A. Al-Betar, S. Naim, S.N. Makhadmeh, Z.A.A. Alyasseri "Link-based multi-verse optimizer for text documents clustering." *Applied Soft Computing* 87 (2020): 106002.
- [3] Kilpeläinen, Jaakko "Automating knowledge work of service desk: Machine learning model for software robot." (2019).
- [4] S.P. Paramesh, K.S. Shreedhara "Automated it service desk systems using machine learning techniques." *Data Analytics and Learning*. Springer, Singapore, 2019. 331-346.
- [5] M. Younas, K. Wakil, M. Arif, A. Mustafa "An automated approach for identification of non-functional requirements using Word2Vec model." *Int. J. Adv. Comput. Sci. Appl* 10 (2019).
- [6] Christopher D. Manning "The Stanford CoreNLP Natural Language Processing Toolkit" *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (2014): 55-60.
- [7] M. Dli "Application of Fuzzy Decision Trees for Rubricating Unstructured Electronic Text Documents" *Proceedings of the IS-2019 Conference* (2019): 108-118.
- [8] Hu, Kai "Understanding the topic evolution of scientific literatures like an evolving city: Using Google Word2Vec model and spatial autocorrelation analysis." *Information Processing & Management* 56.4 (2019): 1185-1203.
- [9] Matt J. Kusner "From word embeddings to document distances" *ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning* (2015): 957-966.